

EXHIBIT 11

**IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF PENNSYLVANIA**

DUANE E. NORMAN, SR.	:	
<i>on behalf of himself and</i>	:	
<i>all similarly situated individuals,</i>	:	
	:	
Plaintiff,	:	
	:	
v.	:	Civil Action No. 18-cv-05225-GAM
	:	
TRANS UNION, LLC,	:	
	:	
Defendant.	:	

REPLY REPORT OF JONATHAN JAFFE

*in response to the March 11th, 2022 report of David Lasater
and the March 11th, 2022 disclosure of Ahunya Tilghman*

APRIL 15th, 2022

I, Jonathan Jaffe, declare as follows:

1. Plaintiff's counsel asked me to prepare this reply report in response to David Lasater's expert report dated March 11th, 2022.¹
2. My main opinions are as follows:

REPLY OPINIONS

3. Dr. Lasater uses a flawed random sampling methodology by selecting sampled documents from chronological segments.
4. Even if one were to accept Dr. Lasater's sampling results, those results do not support Dr. Lasater's conclusions.
5. Dr. Lasater's threshold of 60 miles to presume or suggest that a consumer did not personally mail the letter reflects an incomplete understanding of how the United States Postal Service postmarks letters.
6. Consumers have access to a wide variety of websites providing dispute templates.
7. Dr. Lasater's substantive categories #5-16 that Dr. Lasater purports would collectively require 73,000 hours of review can be determined within minutes by a search of the extracted text of the letters with the ability for a reviewer to confirm the validity of the categorization by only being presented with the contextual portions around the area(s) used to determine the category(ies).

¹ I produced my initial report on March 11th. Mr. Lasater's report, produced the same day, does not address the opinions I expressed in my March 11th report.

8. I have demonstrated the ability and feasibility to automatically categorize Dr. Lasater's substantive categories #5-16.
9. I have demonstrated an alternative method by which to objectively determine letters with indicia that they were sent by credit repair organizations "CROs."
10. The random sample of 75 letters pulled by Ms. Tilghman does not apply to the produced set of 440,073 letters.
11. I have demonstrated the ability objectively to identify letters reflecting the exemplars cited by Ms. Tilghman in her disclosure.

**OPINION: DR. LASATER USES A FLAWED RANDOM SAMPLING METHODOLOGY
BY SELECTING SAMPLED DOCUMENTS FROM CHRONOLOGICAL SEGMENTS.**

12. Although Dr. Lasater correctly identified the number of samples he would need to select from the population of produced letters in order to have a 95% confidence level with certain levels of margin of error,² he biased the selection of his sample by dividing the letters into 16 chronological segments and selecting 25 random letters from each segment.
13. Dr. Lasater describes this process in paragraph 26 of his expert report.

² Although he does not attribute it, Dr. Lasater uses the same Cochran Formula for binary occurrence of a characteristic in a large population as I employed in my sampling of letters in order to determine whether they had indicia of postage: Cochran, W. G. (1977). Sampling techniques (3rd ed.). New York: John Wiley & Sons. Cochran's formula is as follows: $n = Z^2 p(q) / e^2$ where n is the resulting sample size, p is the expected result, q = 1 - p, Z is the Z table value for the confidence level and e is the margin of error. The Z table value of 95% confidence level is 1.96. For smaller populations under ~2,000, Cochran's formula may be adjusted to decrease the number. Dr. Lasater rounds up 384.16 to 385 when determining the appropriate sample size for a 95% confidence level with a 5% margin of error. I discuss that the effective difference is 5 letters in a population of this size.

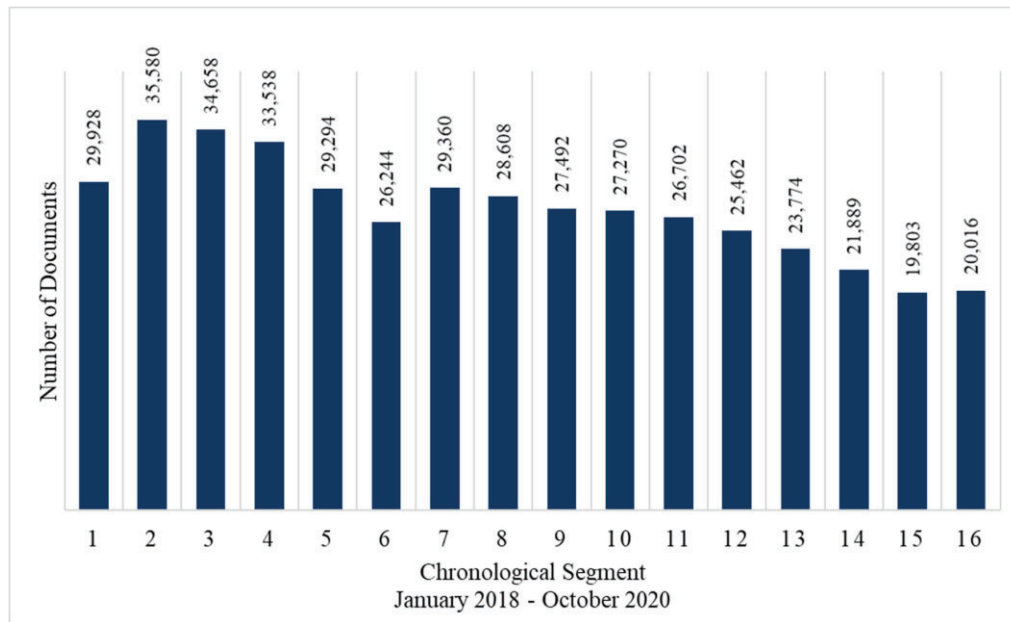


Figure 1- Documents in Class Period (34 Months) Divided into 16 Chronological Segments (Figure 2 in Dr. Lasater's Report)

14. When you select a random sample, that sample is only a representative sample of the population from which you selected it. In effect, Dr. Lasater set up 16 random samples.
15. He introduced the bias he was intending to cure. Namely, what he described in paragraph 25 of his report: “in order to assure **unbiased** sample coverage across the entire 34-month period of the population.”
16. To illustrate why, let’s assume that Dr. Lasater is asking his reviewers to review his initial sample of 400 letters for one of his substantive categories, “Letter Type 3”.
17. Let’s posit, for the sake of my illustration, that “Letter Type 3” was 50% of the letters received by Trans Union during periods 2 and 3³ as highlighted below:

³ A given letter type could be concentrated in a particular time period for any number of reasons, for example: (1) there was a specific instruction given to consumers by Trans Union at that time, (2) there was a popular YouTube, government, or other source at that time with a specific recommendation, (3) there was a popular template employed and available at that time later replaced or removed.

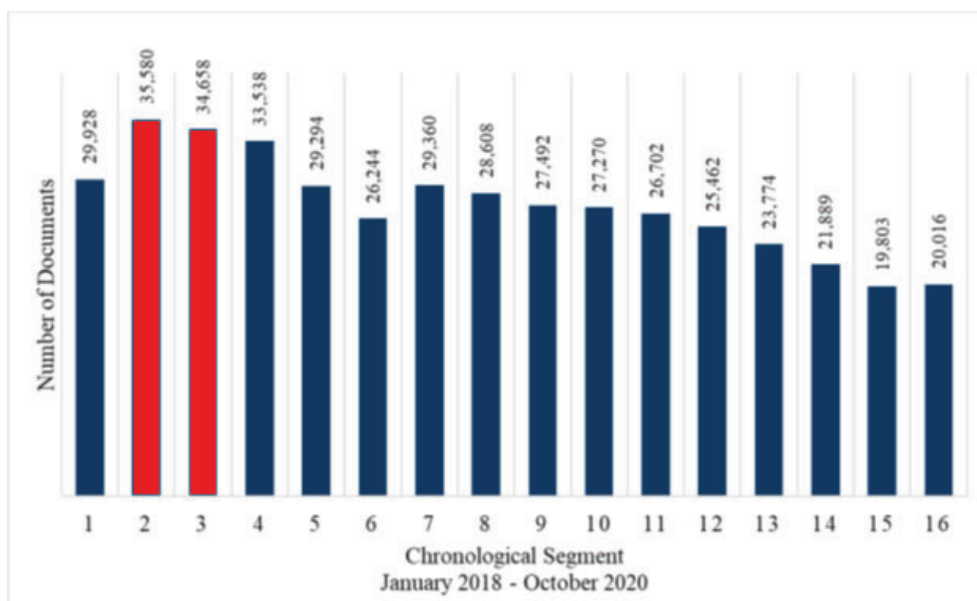


Figure 2 - red bars highlight a possible concentration of letters from a certain template in a particular time period

18. If “Letter Type 3” was 50% of the letters received in periods 2 & 3, then the total number of letters of “Letter Type 3” would be approximately 35,000.
19. To keep the numbers simple, let’s reduce our total population of letters to 350,000. (There are actually 440,073 letters in the population.)
20. That yields 10% of the population is “Letter Type 3.”
21. However, because these letters are not randomly distributed temporally, Dr. Lasater’s random selection would be guaranteed to not have any letters of “Letter Type 3” from periods 1 & 4-16. In other words, each time he pulled 25 letters from those periods, he would find zero of “Letter Type 3.”
22. When he pulled from periods 2 and 3, he would have a 50% chance of getting “Letter Type 3.” Thus, his method would identify 25 letters out of 400 as “Letter Type 3” instead of the actual 40 out of 400.
23. Now, Dr. Lasater does point out that the second sample he took largely confirmed the letter categorization of the first.
24. I performed a Paired T Test on the percentage results of his first and second samples identified in Table 6 and confirmed that there was no statistical difference.⁴
25. However, this is a classic case of **confirmation bias**.⁵ In my hypothetical example of “Letter Type 3” above, a repeat sampling (even a slight increase to 26 sampled per time period) would still yield the same expected error.
26. In my hypothetical example, the error introduced would have translated into missing nearly 40% of “Letter Type 3.”

⁴ The two-tailed P value from comparing the % of sample columns for each group resulted in 0.3153, which is < 0.5, and did not indicate a statistically significant difference.

⁵ Confirmation bias was introduced by assuming that each category was uniformly distributed across the periods.

27. I will note that Dr. Lasater introduced additional bias during his second sample by excluding documents returned in his first sample as he described in paragraph 39 of his expert report.
28. The net effect is that Dr. Lasater's random sampling was only representative for the time period from which the letters were selected.
29. Thus, for each of his 16 periods, he chose 25 letters from ~20-30,000. Using Cochran's formula to determine the margin of error:

$$\sqrt{(1.96^2 \times .5^2 \div 25)} =$$

$$0.196$$

30. Adding sampled results from independent⁶ populations with the same error yields the same error. So, his actual margin of error for his categorizations should be 20% not 5%. If he had wished to account for these 16 periods and maintain a 5% margin of error, then he would have had to have taken 384 letters from each of the 16 periods or a total of 6,144 samples.

OPINION: EVEN IF ONE WERE TO ACCEPT DR. LASATER'S SAMPLING RESULTS, THOSE RESULTS DO NOT SUPPORT DR. LASATER'S CONCLUSIONS.

31. Dr. Lasater outlines 4 opinions in paragraph 13 of his expert report.
32. For Opinion 1, he concludes: "the population of letters present a wide variety of complaints, assertions, questions, and demands."

⁶ Since the time periods are mutually exclusive, these are independent populations.

33. None of his statistical analyses examine the variability of “complaints, assertions, questions, and demands.” As such, his Opinion 1 is unsupported.
34. For Opinion 2, he concludes: “between approximately 177,397 (40.4%) and 206,940 (47.1%) of the 439,618 letters do not complain directly about the accuracy of an inquiry on a consumer’s Trans Union file.”
35. Dr. Lasater’s basis for Opinion 2 is described in paragraph 49 of his report, which is that based upon discussion with Trans Union’s counsel 5 of the 15 categories appeared to be directly related to the accuracy of an inquiry.
36. But, Dr. Lasater did not have his reviewers indicate whether the sampled letters were directly related to the accuracy of an inquiry despite the fact that letters not falling into one of the 5 categories were not categorically known not to have directly related.
37. Thus, his statistical methodology does not support his conclusion for his Opinion 2.
38. For Opinion 3, he concludes: “between approximately 129,225 (29.4%) and 157,185 (35.8%) of the 439,618 letters sent to Trans Union are from addresses at least 60 miles distant from those identified in the letters for the consumers purportedly submitting the letters.”
39. On page 22 of Exhibit 6.2 to Dr. Lasater’s report, Dr. Lasater provided the following instruction as part of the reviewer training:

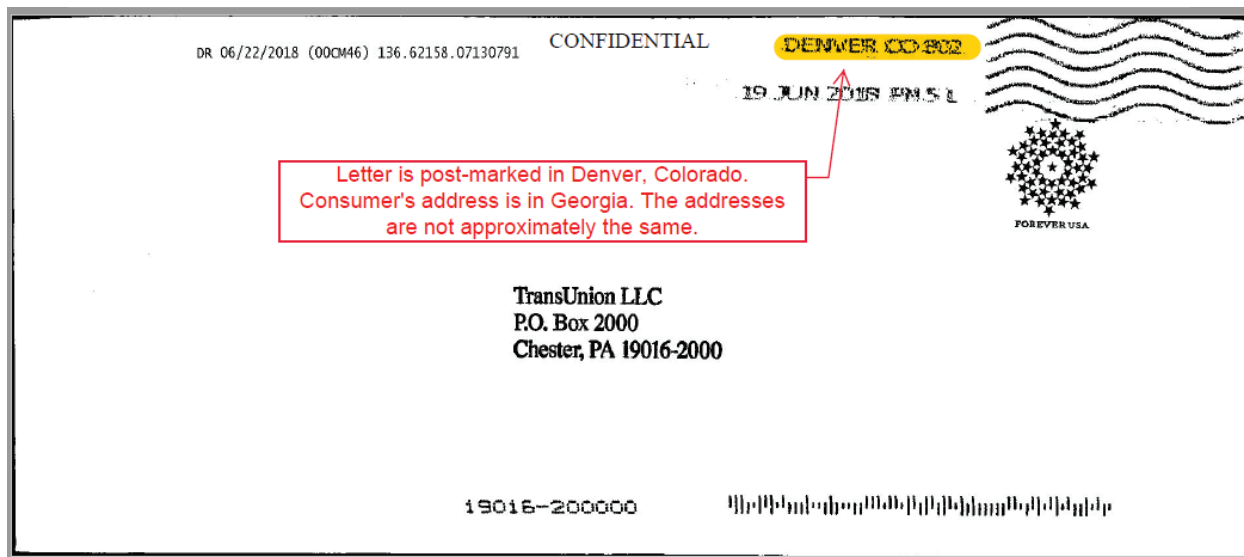


Figure 3 - instruction to reviewers to examine postmarks

40. Dr. Lasater does not instruct his reviewers to look for the consumer's signature when looking at the postmark. The signature to the letter above is on page 20 of Exhibit 6.2 in Dr. Lasater's reviewer training instructions:

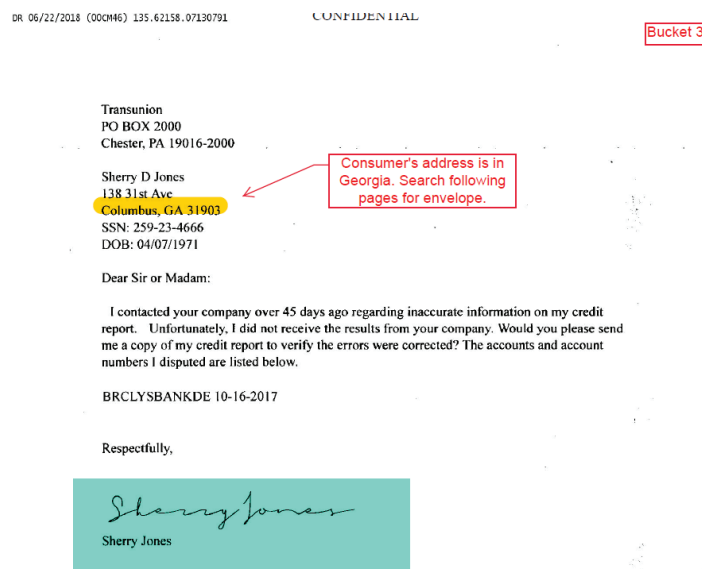


Figure 4 - the signature of the consumer is present, but not highlighted on the original instruction to the reviewers

41. As I elaborate further below, USPS postmarks are most frequently indications of the Processing Distribution Center, not the local post office. You can tell that the example postmark above is not a “local” postmark because it does not contain the full name of the post office followed by the two-letter state abbreviation and the zip code. Dr. Lasater’s training manual does not distinguish between “local” postmarks and postmarks applied at the Processing Distribution Center.
42. Thus, since reviewers were not taking this distinction into account, the statistical analysis based upon the review of the sample letters does not support Dr. Lasater’s Opinion 3.
43. For Opinion 4, Dr. Lasater concludes that his observation that many letters used templates “creates uncertainty that the complaints, assertions, questions, or demands in the letters reflect bona fide points of view or perspectives of the consumers whose names are associated with the letters.”
44. As dispute letter templates are widely available on the Internet, including at least some of the 16 templates identified by Dr. Lasater, the statistics do not support his conclusion. Notably, Dr. Lasater provides no confidence interval for Opinion 4.

OPINION: DR. LASATER’S THRESHOLD OF 60 MILES TO PRESUME OR SUGGEST THAT A CONSUMER DID NOT PERSONALLY MAIL THE LETTER REFLECTS AN INCOMPLETE UNDERSTANDING OF HOW THE UNITED STATES POSTAL SERVICE POSTMARKS LETTERS.

45. Chapter 1, section 1.3, of the USPS Handbook on Area Mail Processing Guidelines describes the application and various methods of imprinting postmarks:

1-1.3 Postmarks

A postmark is an official Postal Service™ imprint applied in black ink on the address side of a stamped mailpiece. A postmark indicates the location and date the Postal Service accepted custody of a mailpiece, and it cancels affixed postage. Since 1979, the Postal Service's *Postal Operations Manual* (POM) has provided standards for postmarks applied to single-piece First-Class Mail®. Letters and flats that need to be postmarked come from carrier pick-up, collection boxes, retail counters, or lobby drop boxes. Postmarks are not required for mailings bearing a permit, meter, or precanceled stamp for postage, nor to pieces with an indicia applied by various postage evidencing systems.

The postmarking process uses the following three basic methods of imprinting:

1. **Automated:** Advanced facer canceller systems used by processing distribution centers cancel letters quickly. These machines are equipped with biohazard detection systems so letters postmarked by automation benefit from added safety measures.
2. **Mechanized:** A variety of older devices apply postmarks to flat-size mailpieces and to philatelic pieces.
3. **Manual:** Hand-stamp devices are used by Postal Service employees for local cancellation or philatelic requests.

A "local" postmark shows the full name of the Post Office, a two-letter state abbreviation, ZIP Code™, and date of mailing. Because the Postal Service is sensitive to the importance some customers place upon these postmarks, each Post Office is required to make a local postmark available. Lobby drops should be designated for this purpose with clear signage signifying its use.

Figure 5 - Chapter 1, § 1.3, USPS Handbook PO-408 - Area Mail Processing Guidelines, retrieved 2022-04-06 from https://about.usps.com/handbooks/po408/ch1_003.htm (highlighted)⁷

46. Note that “local” postmarks show the full name of the Post Office. The examples that Dr. Lasater provided his reviewers were not local postmarks imprinted at the local post office, but postmarks imprinted at processing distribution centers.
47. Mail can be sent more than 60 miles from the site of local collection to the nearest Processing Distribution Center.

⁷ Attached hereto as **Exhibit E - 1-1.3 Postmarks.pdf**.

48. As of 2011, the USPS had 251 Processing and Distribution Centers (“P&DCs”) across the United States. That was down from 269 in 2005.⁸



FACT SHEET PROCESSING FACILITIES

The Postal Service has nine different types of processing facilities:

Processing and Distribution Centers — P&DCs process and dispatch mail coming in from Post Offices and collection boxes in a specific geographic location.

Customer Service Facilities — CSFs are Post Office stations or branches that provide

How many of each kind of processing facility are there? Has the number changed in the past five years?

PROCESSING FACILITY TYPE		2011	2010	2009	2008	2007	2006	2005
Processing and Distribution Centers	P&DC	251	260	268	269	269	269	269
Customer Service Facilities	CSF	115	164	195	195	195	195	195
Network Distribution Centers	NDC	21	21	21	21	21	21	21
Logistics and Distribution Centers	LDC	10	13	14	14	14	11	11
Annexes	—	46	51	61	64	66	66	66
Surface Transfer Centers	STC	10	11	20	20	14	17	14
Air Mail Centers	AMC	1	1	12	20	29	77	79
Remote Encoding Centers	REC	2	2	3	6	10	12	15
International Service Centers	ISC	5	5	5	5	5	5	5
Total Processing Facilities		461	528	599	614	623	673	675

Figure 6 - USPS Electronic Press Kit Publication on Processing Facilities retrieved 2022-04-06 from https://about.usps.com/news/electronic-press-kits/our-future-network/processing_facility_types.pdf (condensed and highlighted)⁹

49. The United States has a total land area of 3.53 million square miles.¹⁰
50. There are 351 cities in the United States with over 100,000 individuals and over 19,502 incorporated places.

⁸ In recent years there has been discussion about the consolidation and closure of a further 18 facilities. <https://www.nbcnews.com/news/us-news/u-s-postal-service-consolidate-18-facilities-leading-concerns-over-n1265748>

⁹ Attached hereto as **Exhibit F** - [processing_facility_types.pdf](#).

¹⁰ <https://www.census.gov/geographies/reference-files/2010/geo/state-area.html>

51. Even with USPS P&DCs located in major population areas, a considerable percentage of the US population would be more than 60 miles from a USPS P&DC.
52. With under 300 USPS P&DCs, it is clear that Dr. Lasater's 60-mile threshold is arbitrary and unsupported for mail with post marks imprinted at a USPS P&DC.
53. Furthermore, Dr. Lasater does not attempt to quantify the number of individuals *in this population* who send mail for legitimate reasons at a distance from their address, whether because they are sending it in the vicinity of their workplace, sending it from a location at which they are temporarily due to travel or other reasons, or having someone help them print and send the letter.

OPINION - CONSUMERS HAVE ACCESS TO A WIDE VARIETY OF WEBSITES PROVIDING DISPUTE TEMPLATES.

54. A quick internet search yielded easily accessible inquiry dispute templates available on the Internet at no cost:
 - a. <https://www.coursehero.com/file/31903263/Experian-Inquiry-Removal-Letterdocx/>
 - b. <https://www.pdfFiller.com/486897324--REQUEST-TO-RE-INVESTIGATE-AND-DELETE-THE-UNAUTHORIZED-INQUIRY-TRANSUNION-iDisputeorg->
 - c. <https://www.lexingtonlaw.com/blog/negative-items/credit-inquiry-removal.html>
 - d. <https://www.imaxcredit.com/credit-dispute-letters-credit-bureau/>
 - e. <https://www.uslegalforms.com/us/US-01600BG.htm>

- f. <https://community.debtcc.com/letters/sample13.html>
- g. <https://www.youtube.com/watch?v=AGUZ5paqvCQ>

55. Many of these templates are substantially similar to the templates identified by Dr. Lasater:

07/16/2016

Your Name, as on credit
Your Address, as on credit
City, ST 123456

Experian
PO Box 9701
Allen, TX 75013
REF: YOUR-SSN

Dear Experian:

I am very concerned regarding the serious possibility that you have not followed the permissible purpose provisions of the federal FCRA. A credit inquiry for CREDITOR NAME was placed within my credit file in July, 2016. The Act specifies that you must document your reasons for including such items

Figure 7 - Experian Inquiry Removal Letter from Course Hero link above

- 56. Many of the instructions accompanying these templates list the mailing addresses of the three major credit bureaus including the Defendant Trans Union.
- 57. Given the accessibility of these templates, the use of a template by itself does not support Dr. Lasater's speculation in paragraph 51 of his report, based upon his discussions with Trans Union counsel, that an agent sent the letter.
- 58. There are no other indicia besides USPS P&DC postmarks and widely available letter templates that Dr. Lasater uses as his basis for his second overarching topic described in paragraph 9 of his report to support his conclusion that consumers did not send these dispute letters.

59. There are indicia, even in Dr. Lasater's examples to his reviewers, that the consumer endorsed the letter with the consumer's signature and included the consumer's return address, not the return address of a purported agent.

OPINION - DR. LASATER'S SUBSTANTIVE CATEGORIES #5-16 THAT DR. LASATER PURPORTS WOULD COLLECTIVELY REQUIRE 73,000 HOURS OF REVIEW CAN BE DETERMINED WITHIN MINUTES BY A SEARCH OF THE EXTRACTED TEXT OF THE LETTERS WITH THE ABILITY FOR A REVIEWER TO CONFIRM THE VALIDITY OF THE CATEGORIZATION BY ONLY BEING PRESENTED WITH THE CONTEXTUAL PORTIONS AROUND THE AREA(S) USED TO DETERMINE THE CATEGORY(IES).

60. In paragraph 21 of Dr. Lasater's expert report, he estimates that it would take reviewers more than 73,000 hours of linear review time to address "whether the letters, taken as an entirety, reflect a singular complaint, assertion, question or demand of Trans Union."
61. Thus, by implication, he implies that every single letter would need to be read in its entirety in order to determine whether or not the letter reflected a true dispute.
62. Of course, Trans Union has represented that these are all indeed dispute letters.
63. Setting aside that these were all letters sent to Trans Union and tracked by Trans Union as disputes, the substantive categorizations #5-13 outlined by Dr. Lasater in Table 2 on page 13 of his report can be determined in a less subjective manner through a series of queries against their text.

64. During the process of identifying the text matching the categorization, I am able to extract the contextual text around what was identified. This contextual text allows a reviewer to perform due diligence upon the resulting automatic categorization for any further refinement to the terms being searched.
65. I took a random sample of 10,000 letters¹¹ (from the 440,073 letters produced by Trans Union).
66. I identified 13 categories of terms consistent with dispute letters:
- a. dispute
 - b. not mine
 - c. not or never applied
 - d. unauthorized
 - e. unauthorized inquiry
 - f. permission or permissible purpose
 - g. not authorized
 - h. inquir(ies)
 - i. remove or deleted or erase
 - j. inaccurate
 - k. legal terms
 - l. do not recall
 - m. consent

¹¹ I took two independent samples of 10,000 letters. The first sample helped determine the categories I used against the second random sample. I took a second sample to avoid any confirmation bias after adjusting my queries. Herein, where I refer to the sample of 10,000 letters, I am uniformly referring to the random second sample of 10,000 letters I took. The sample was taken randomly using the sample_n function in R.

67. I used a regular expression¹² engine to further refine the searches for each of these categories. The regular expressions (corresponding to each category) I employed were:

- a. account.{140}disput|dispute.{140}account|wish.{140}dispute|report dispute
form|i dispute|writ.{140}disput|disput.{140}item|procedure in case of disput|my
disput|reinvestigations? of disput|dispute center|dispute.{110}following
- b. not mine|not me
- c. i.{140}not appl|i.{140}never appl
- d. unauthorized
- e. unauthorized.{140}inquir
- f. permission|permissible purpose
- g. not authoriz|never authoriz
- h. inquire(?!ing)
- i. please remov|permanent
delet|erase|remov.{120}inqu|take.{120}off.{120}report|remove the following
- j. inaccura|not accura
- k. law enforcement|and federal law requires|procedure in case of disputed
accuracy|complaint and affidavit
- l. don't recall|do not recall
- m. didn't consent|did not consent|validate my consent|consent to
pull|demonstrate.{130}hard|provide proof|inquir.{130}remov

¹² Regular expressions search the exact text instead of an index. The | or pipe symbol indicates a logical OR. The . or period indicates any character. The {} braces denote the number of repeated characters. The (?!) question mark followed by exclamation point is a negative lookahead, meaning does not have these characters following. The ? questions mark means a character was optional. Unlike an index, regular expressions can search for any word including stop words that are generally removed from indices, such as **not, in, of, I, the, and, to, et cetera.**

68. I ran each of the 13 searches against the extracted/OCR'd text of all 10,000 letters.¹³
69. The 13 searches completed in 93.89 seconds. That included extracting the contextual text surrounding each term, which I set at 40 characters preceding and 40 characters following the first¹⁴ match.
70. I exported the contextual results for each letter as the spreadsheet hereto attached as **Exhibit A - 10K sample No. 2 terms with context - Norman v Trans Union - SUBJECT TO PROTECTIVE ORDER.xlsx**.
71. The entire process from generating the sample, running the searches, and extracting the results took under 5 minutes.
72. 10,000 Letters is ~2.27% of the number of produced letters. While the time could vary depending on the length of the letter text, I project it would take ~250 minutes or a little over 4 hours of compute time to perform the same exercise on the text of all 440,073 produced letters.
73. Since I pull the contextual text, reviewers confirming the relevance of a particular term would not have to spend time reading every letter, but only the extracted text.
74. I augmented the results with a count of how many categories in which terms were found. The results were:
- a. Letters with no categories with matching terms: 329 (3.3%)
 - b. 1 or more categories with matching terms: 9,671 (96.7%)
 - c. 2 or more categories with matching terms: 9,191 (91.9%)
 - d. 3 or more categories with matching terms: 7,042 (70.4%)

¹³ Each regular expression search was run case independent (treating UPPERCASE and lowercase and Mixed Case text the same).

¹⁴ For review purposes, it's most efficient to pull the first matched hit for examination.

e. 4 or more categories with matching terms: 3,571 (35.7%)

75. The R code for this analysis is hereto attached as **Exhibit D** - data analysis and source code\rebuttal report analysis\run terms against large sample (round 2).R.
76. Using Cochran's formula for binomial distributions, we can compute the margin of error for a 10K sample. Since our sample is large, instead of a 95% confidence level, we will use a 99% confidence level, which has a Z table value of 2.58.

$$e = \sqrt{\frac{Z^2(p)(q)}{10,000}} = \sqrt{\frac{2.58^2(0.5^2)}{10,000}} = \sqrt{\frac{1.6641}{10,000}} = \sim 0.0129$$

77. Thus, with a 99% level of confidence, $91.9\% \pm 0.13\%$ of the produced letters would have language matching 2 or more terms out of the 13 categories I searched using the regular expressions specified above.

OPINION - I HAVE DEMONSTRATED THE ABILITY AND FEASIBILITY TO AUTOMATICALLY CATEGORIZE DR. LASATER'S SUBSTANTIVE CATEGORIES #5-16.

78. By performing the aforementioned exercises to identify the letters with the language as requested by Plaintiff's Counsel, I have demonstrated the ability and feasibility to automatically¹⁵ categorize Dr. Lasater's substantive categories #5-13 and any such additional categories that are based upon similar textual analysis.
79. Note that any such analysis will require a relatively small amount of additional manual review for letters that were handwritten or poorly OCR'd.

¹⁵ Excluding handwritten and poorly OCR'd letters which constitute a small minority of the letter population.

80. Should the Court or the parties determine a set of categories, and a series of exemplar letters or search terms, to be included in the Class **or** excluded from the Class **or** merely needing to be further inspected before a decision certain is made as to inclusion or exclusion, I would be able to construct and run queries to identify such letters.

OPINION: I HAVE DEMONSTRATED AN ALTERNATIVE METHOD BY WHICH TO OBJECTIVELY DETERMINE LETTERS WITH INDICIA THAT THEY WERE SENT BY CREDIT REPAIR ORGANIZATIONS “CROs.”

81. Dr. Lasater used the Postmark and templates as indicia that letters in the production were sent by a Credit Repair Organization “CRO.”
82. His methodology employed a labor-intensive review which for reasons I stated *supra* was flawed.
83. I have come up with an alternative way to identify letters more likely to have been sent by a CRO using indicia of bulk rate postage.
84. Attached hereto as Exhibit G - What is Commercial Mail__ Postal Explorer.pdf is indication that in order to qualify for bulk rate postage, the minimum number of pieces being mailed must be 500.
85. Examining a large sample of PDF letters and extracted/OCR’d text, I identified eight categories of indicia for bulk mailing:
- a. presorted
 - b. pitney bowes
 - c. neopost
 - d. zip 32801

- e. stamps endicia
 - f. letterstream
 - g. privacy guard
 - h. see important information enclosed
86. Term (a) included all indicia on the postage mark as further described in the regular expressions below indicating a bulk postage rate.
87. The terms (b)-(f) were indicative of a bulk rate stamp by a third-party vendor.
88. Terms (g) and (h) were indicative of language common to a bulk mailing.
89. As with the aforementioned dispute letter search terms, I employed regular expressions for these “CRO” searches corresponding to the categories above. These were:
- a. `presorted| pres.{3}ed f|prsrtd|postage.{1,20}paid.{1,30}permit|pre-sorted|permit no\\.?.? 1`
 - b. `p.tney b`
 - c. `neopost`
 - d. `zip 32801`
 - e. stamps endicia
 - f. letterstream
 - g. privacy guard
 - h. see important information enclosed
90. I ran these regular expressions against the aforementioned random sample of 10,000 letters.
91. The processing time to run the above queries against the extracted/OCR'd text of the 10,000 sampled letters was 6.89 seconds.

92. I exported the results of these searches (which included the contextual text) attached hereto as **Exhibit B** - 10K sample CRO terms with context - Norman v Trans Union - SUBJECT TO PROTECTIVE ORDER.xlsx.
93. I augmented this spreadsheet with a count of how many matching categories for each document. The results were:
- a. Letters with no categories with CRO matching terms: 9,549 (95.5%)
 - b. 1 or more categories with matching terms: 406 (4.1%)
 - c. 2 or more categories with matching terms: 133 (1.3%)
 - d. 3 or more categories with matching terms: 32 (0.3%)
94. Thus, it is my opinion, within a 99% confidence level, that $4.1\% \pm 0.13\%$ of the letters in the production have indicia associated with a bulk mailing suggestive of a CRO.
95. The R code for this analysis is hereto attached as **Exhibit D** - data analysis and source code\rebuttal report analysis\run CRO terms against large sample (round 2).R.
96. Again, manual reviewers confirming this would not have to read the text of every letter, but only the contextual text around the search for the ~4% of letters where text matching this category was found.
97. Should the Court or the parties determine a set of categories indicative of letters sent by a CRO, and a series of exemplar letters or search terms, either to be included in the Class **or** excluded from the Class **or** merely needing to be further inspected before a decision certain is made as to inclusion or exclusion, I would be able to construct and run queries to identify such letters.

OPINION: THE RANDOM SAMPLE OF 75 LETTERS PULLED BY MS. TILGHMAN DOES NOT APPLY TO THE PRODUCED SET OF 440,073 LETTERS.

98. In paragraph 7(b) of the March 11th disclosure, Trans Union describes how Ms. Tilghman arrived at her random sample of 75 letters:

“To assemble her random sample, Ms. Tilghman prepared a report identifying “Fraud and Non-Fraud Inquiry Deletes”—i.e., a report that identifies instances when Trans Union has removed an inquiry from a consumer’s credit file, as well as the general reason for the removal, which was similar to TU02032258—and randomly selected 75 consumers whose inquiries were, according to the report, deleted “per documentation,” i.e., instances when an inquiry was removed based on a purported letter from the creditor (or other documentation) sent by the consumer.”

99. Since her random sample was not drawn from the population of the 440,073 letters, but from instances where Trans Union had removed an inquiry from a consumer’s credit file for reasons that were similar to TU02032258 **and** whose inquiries were deleted “per documentation,” this random sample does not apply to the produced set, and any conclusions drawn by Ms. Tilghman about her sample only apply to the universe of letters from which she sampled, not the population of letters produced in this case.

OPINION: I HAVE DEMONSTRATED THE ABILITY OBJECTIVELY TO IDENTIFY LETTERS REFLECTING THE EXEMPLARS CITED BY MS. TILGHMAN IN HER DISCLOSURE.

100. Even though Ms. Tilghman’s conclusions about her random sample cannot be applied to the population of letters produced in this case, I did examine the letters cited in her disclosure as exemplars of containing both fraudulent and non-fraudulent letters from

purported third parties: TU00252792, TU00360392, TU00425364, TU01942355, TU01942875, TU01949373.

101. Upon examination, I came up with the following categories to try to identify similar letters in my 10,000 sample:

- a. apologize/apologies
- b. inconvenience
- c. our mistake
- d. regret
- e. failure on our part

102. I constructed regular expressions as before corresponding to each category:

- a. apologi
- b. inconvenience
- c. mistake.{1,10} our| our.{1,10}mistake
- d. regret
- e. failure.{1,20} our| our.{1,20}failure

103. I executed this search against the 10K sample.

104. The search completed in 4.98 seconds.

105. The results are attached hereto as **Exhibit C** - 10K sample Tilghman disclosure terms with context - Norman v Trans Union - SUBJECT TO PROTECTIVE ORDER.xlsx.

106. There were no letters that had terms matching the “our mistake” or “failure on our part” categories.

107. Upon further examination of the other terms, there were false positives for the standalone: “apologize/apologies” “inconvenience” “regret”

108. After re-examining Ms. Tilghman's examples, I noted that in all but one of the letters she identified as fraudulent, the phrase "apologize for any inconvenience" or some slight variation on that was present.
109. Accordingly, filtering for letters that had apologize/apologies and inconvenience yielded 14 letters, or 0.14% of the 10,000 sampled.
110. The R code for this analysis is hereto attached as **Exhibit D** - data analysis and source code\rebuttal report analysis\run suspect fraud terms against large sample (round 2).R.
111. It is my opinion that using the above process and a manual review of the contextual text surrounding those results, letters similar to those Ms. Tilghman identified could be ascertained.
112. It is my opinion, within a 99% level of confidence, $0.14\% \pm 0.13\%$ of the letters in the production have language matching Ms. Tilghman's disclosure exemplar letters.
113. Should the Court or the parties determine a set of categories indicative of fraudulent letters in the same vein as those identified by Ms. Tilghman, and a series of exemplar letters or search terms, either to be included in the Class **or** excluded from the Class **or** merely needing to be further inspected before a decision certain is made as to inclusion or exclusion, I would be able to construct and run queries to identify such letters.

MATERIALS REVIEWED AND CONSIDERED FOR THIS REPLY REPORT

114. All prior materials cited in my Expert Report dated March 11th, 2022.
115. David Lasater Expert Report dated March 11th, 2022.
116. Ahunya Tilghman Expert Disclosure dated March 11th, 2022.
117. Ahunya Tilghman Expert Disclosure materials cited.

118. USPS Postal Operations Handbook, specifically Chapter 1, § 1.3, USPS Handbook
PO-408 - Area Mail Processing Guidelines:
https://about.usps.com/handbooks/po408/ch1_003.htm
119. USPS Electronic Press Kit Publication on Processing Facilities:
https://about.usps.com/news/electronic-press-kits/our-future-network/processing_facility_types.pdf
120. USPS Bulk Mail Minimum Quantities:
<https://pe.usps.com/businessmail101?ViewName=WhatIsBulkMail>
121. US Census Data reflecting Land Area in Sq. Miles of the United States:
<https://www.census.gov/geographies/reference-files/2010/geo/state-area.html>
122. PDFs of letters: TU00648822, TU00693089, TU00734132, TU00761409, TU00872889, TU00878888, TU00923494, TU00926103, TU00974193, TU00989746, TU00002259, TU00112820, TU00118051, TU00217559, TU00218345, TU00220629, TU00270514, TU00349208, TU00349226, TU00382948, TU00473790, TU00630178, TU00643637, TU00648316, TU00526475, TU00634224, TU00825722, TU00872000, TU00894083, TU00986633, TU01025673, TU00252792, TU00360392, TU00425364, TU01942355, TU01942875, TU01949373, TU01949626, TU01949804, TU01950361, TU01950367, TU01964605, TU01971317, TU01995357, TU01998707, TU01999175, TU02002570, TU02002996, TU02003762, TU02013307, TU02014894, TU02015144, TU00002787, TU00005883, TU00007508, TU00008308, TU00012929, TU00013044, TU00015926, TU00016028, TU00017820, TU00018861, TU00020475, TU00021478, TU00022793, TU00031402, TU00031611, TU00033865, TU00037248, TU00039904, TU00051572, TU00051578, TU00051717, TU00066776, TU00084715, TU00085925, TU00091188,

TU00096040, TU00096682, TU00099171, TU00099438, TU00101632, TU00104239, TU00106945, TU00108085, TU00111299, TU00113342, TU00115556, TU00121748, TU00125088, TU00128489, TU00134528, TU00134763, TU00135109, TU00141735, TU00145422, TU00152814, TU00158401, TU00160589, TU00162271, TU00165888, TU00167548, TU00168344, TU00175319, TU00183077, TU00183180, TU00183935, TU00184731, TU00189184, TU00193100, TU00196973, TU00199457, TU00203345, TU00208875, TU00214246, TU00215018, TU00215384, TU00223162, TU00226279, TU00227266, TU00227508, TU00231511, TU00239377, TU00242766, TU00248707, TU00253607, TU00256364, TU00259524, TU00259941, TU00262506, TU00264079, TU00270597, TU00275613, TU00277082, TU00279298, TU00284518, TU00288958, TU00291898, TU00294312, TU00298647, TU00302549, TU00338723, TU00338732, TU00339276, TU00353669, TU00360848, TU00364367, TU00366179, TU00367105, TU00371038, TU00375718, TU00379500, TU00383541, TU00387357, TU00392749, TU00395343, TU00398844, TU00414906, TU00418326, TU00425964, TU00430681, TU00436846, TU00444329, TU00452123, TU00455104, TU00457962, TU00460272, TU00471470, TU00477271, TU00484111, TU00492991, TU00500162, TU00507964, TU00513763, TU00520779, TU00531757, TU00549249, TU00552950, TU00554449, TU00557691, TU00558144, TU00567367, TU00570613, TU00579575, TU00585670, TU00588852, TU00589029, TU00590518, TU00591028, TU00593070, TU00606381, TU00614586, TU00625182, TU00626135, TU00628943, TU00633267, TU00634771, TU00634879, TU00637533, TU00653348, TU00657062, TU00663157, TU00666525, TU00671779, TU00674515, TU00679636, TU00692397, TU00701965, TU00705516, TU00706275, TU00714311, TU00715626, TU00720377, TU00744743, TU00759306,

TU00760786, TU00773970, TU00774947, TU00779466, TU00787454, TU00789163, TU00791824, TU00792836, TU00801104, TU00815014, TU00818064, TU00823966, TU00825238, TU00830801, TU00835439, TU00847935, TU00853499, TU00861284, TU00863812, TU00868444, TU00873360, TU00874670, TU00875709, TU00884367, TU00891779, TU00898291, TU00901239, TU00907208, TU00914112, TU00915545, TU00922563, TU00928121, TU00929368, TU00939605, TU00948703, TU00949371, TU00957263, TU00967469, TU00967890, TU00982174, TU00983106, TU00983147, TU00990961, TU00998488, TU01004964, TU01008155, TU01009535, TU01013670, TU01019887, TU01024081, TU01029615, TU01035792, TU01037110, TU01047545, TU01056949, TU01057352, TU01059097, TU01061546, TU01072465, TU01072485, TU01082283, TU01088258, TU01091233, TU01092322, TU01099303, TU01099506, TU01107031, TU01109528, TU01113746, TU01119946, TU01129638, TU01130298, TU01137138, TU01143813, TU01146570, TU01152480, TU01153003, TU01154061, TU01157280, TU01163690, TU01170599, TU01170887, TU01177882, TU01180213, TU01184143, TU01184698, TU01192510, TU01202036, TU01203472, TU01215314, TU01225343, TU01228007, TU01228913, TU01235396, TU01235795, TU01236808, TU01244554, TU01259992, TU01261223, TU01265220, TU01266633, TU01272017, TU01284696, TU01285837, TU01290123, TU01293332, TU01298675, TU01299935, TU01308861, TU01331368, TU01336054, TU01336261, TU01337015, TU01337509, TU01348106, TU01356235, TU01359663, TU01361107, TU01367371, TU01374944, TU01382492, TU01384542, TU01385143, TU01397256, TU01397904, TU01413491, TU01414811, TU01421728, TU01425178, TU01430036, TU01430666, TU01445026, TU01449233, TU01479564, TU01494266, TU01507354, TU01511946, TU01513605,

TU01515777, TU01516667, TU01523225, TU01524127, TU01530509, TU01535027, TU01539193, TU01544005, TU01549831, TU01568238, TU01572600, TU01579035, TU01605563, TU01611445, TU01617567, TU01620616, TU01622055, TU01624053, TU01638649, TU01661593, TU01668935, TU01693566, TU01693650, TU01700067, TU01714335, TU01716322, TU01719580, TU01721320, TU01733763, TU01734125, TU01735680, TU01736664, TU01738993, TU01753284, TU01758241, TU01758885, TU01759957, TU01768247, TU01774596, TU01783586, TU01800694, TU01800921, TU01803561, TU01805720, TU01809200, TU01814182, TU01816198, TU01819556, TU01820299, TU01820972, TU01826047, TU01826207, TU01826302, TU01828175, TU01838630, TU01840536, TU01850593, TU01856706, TU01884348, TU01902928, TU01903730, TU01916023, TU01922595, TU01922900, and TU01934431.

123. Extracted/OCR'd text of the 440,073 letters produced in the litigation.

HARDWARE + SOFTWARE EMPLOYED FOR MY ADDITIONAL ANALYSIS

124. The Hardware Employed for the Data Analysis was as follows:

- a. Dell Alienware Aurora R13
- b. Processor: 12th Gen Intel(R) Core™ i9-12900KF 3.19 GHz
- c. 128 GB RAM

125. Software Employed for the Data Analysis was as follows:

- a. Operating System: Windows 11 v21H2 64bit
- b. Development Environment: RStudio¹⁶ version 2022.02.1 Build 461

¹⁶ RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

- c. Development Language: R version 4.1.3¹⁷
126. R Packages¹⁸ Utilized:
- a. Data Import: vroom¹⁹
 - b. Data Export / Reporting: vroom²⁰, openxlsx²¹
 - c. Data Shaping / Processing: dplyr²², tidyverse²³, stringi²⁴, stringr²⁵, lubridate²⁶, janitor²⁷, openssl²⁸, purrr.²⁹
 - d. Benchmarking: tictoc³⁰, beeper³¹

¹⁷ R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

¹⁸ In the R language it is standard practice to employ various “packages” of open source code.

¹⁹ Jim Hester, Hadley Wickham and Jennifer Bryan (2021). vroom: Read and Write Rectangular Text Data Quickly. R package version 1.5.7. <https://CRAN.R-project.org/package=vroom>

²⁰ Ibid.

²¹ Philipp Schauburger and Alexander Walker (2021). openxlsx: Read, Write and Edit xlsx Files. R package version 4.2.5. <https://CRAN.R-project.org/package=openxlsx>

²² Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. <https://CRAN.R-project.org/package=dplyr>

²³ Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

²⁴ Gagolewski M (2021). “stringi: Fast and portable character string processing in R.” *Journal of Statistical Software*.

²⁵ Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>

²⁶ Garrett Golemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1-25. URL <https://www.jstatsoft.org/v40/i03/>.

²⁷ Sam Firke (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.1.0. <https://CRAN.R-project.org/package=janitor>

²⁸ Jeroen Ooms (2021). openssl: Toolkit for Encryption, Signatures and Certificates Based on OpenSSL. R package version 1.4.6. <https://CRAN.R-project.org/package=openssl>

²⁹ Lionel Henry and Hadley Wickham (2020). purrr: Functional Programming Tools. R package version 0.3.4. <https://CRAN.R-project.org/package=purrr>


³⁰ Sergei Izrailev (2021). tictoc: Functions for Timing R Scripts, as Well as Implementations of Stack and List Structures. R package version 1.0.1. <https://CRAN.R-project.org/package=tictoc>

³¹ Rasmus Bååth (2018). beeper: Easily Play Notification Sounds on any Platform. R package version 1.3. <https://CRAN.R-project.org/package=beeper>

CERTIFICATION

I hereby certify that the foregoing statements made by me are true and correct based on the information available to me at this time. I reserve the right to amend and supplement this report as additional information and evidence is provided to me.

Signed this 15th day of April, 2022.



Digitally signed by Jonathan Jaffe
DN: cn=Jonathan Jaffe,
o=www.its-your-internet.com,
ou=www.its-your-internet.com,
email=jjaffe@its-your-
internet.com, c=US
Date: 2022.04.15 14:40:16 -04'00'
Adobe Acrobat version:
2022.001.20085

JONATHAN JAFFE